Theor Appl Genet (1988) 76:443–448

Primary structure and sequence organization of the 16S–23S spacer in the ribosomal operon of soybean (*Glycine max* L.) chloroplast DNA

G. de Lanversin and D.T.N. Pillay*

Department of Biological Sciences, University of Windsor, Windsor, Ontario N9B 3P4, Canada

Received July 30, 1987; Accepted August 19, 1987
Communicated by G. Wenzel

Summary. The nucleotide sequence of a spacer region between 16S and 23S rRNA genes from soybean chloroplasts has been determined. The spacer region is over 3000 bp long and contains two tRNA genes, coding for $rRNA^{Ile}$ and $tRNA^{Ala}$ which contain intervening sequences of 953 and 811 base pairs respectively. There is a strong homology between the two introns suggesting that they have a common origin. These spacer tRNAs are synthesized as part of a kb precursor molecule containing 16S and 23S rRNA sequences.

Key words: Chloroplast – DNA – Gene mapping – tRNA gene – Nucleotide sequence

Introduction

Chloroplast genome of green plants is a closed circular duplex DNA molecule of 120–180 kb, which contain genes coding for rRNAs, tRNAs and polypeptides. In soybean, a single set of rRNA genes is present per segment of the inverted repeat. The composite rRNA genes in the set are arranged as 5′-16S-spacer-23S-5S-3′ (Singh et al. 1984). The structural genes coding for rRNA species are separated by spacer regions. In the *Z. mays* and the tobacco chloroplast genomes, large spacers of 2400 and 2080 base pairs respectively separate the 16S from the 23S rRNA gene (Bedbrook and Kolodner 1979; Bedbrook et al. 1977; Koch et al. 1981; Takaira and Sugiura 1982). The spacer region in the soybean is over 3000 base pairs. The presence of certain tRNA genes in the 16S–23S spacer regions have been reported in chloroplasts from maize (Koch et al. 1981), spinach (Bohnert et al. 1979), tobacco (Takaira and Sugiura 1982), *Euglena gracilis* (Graf et al. 1980; Orozco et al. 1980), *Chlamydomonas reinhardii* (Malnoe and Rochaix 1978) and *E. coli* (Young et al. 1979; Sekiya and Nishimura 1979; Morgan et al. 1977).

It has been reported that chloroplast 16S and 23S rDNAs in higher plants are separated by much longer spacers (1750–3000 base pairs), compared to those found in *E. coli* 437 bp – (Young et al. 1979) and *E. gracilis* 259 bp (Graf et al. 1980; Orozco et al. 1980). Koch et al. (1981) in maize, Takaira and Sugiura (1982) in tobacco, Graf et al. (1980) and Orozco et al. (1980) in *Euglena* have sequenced the spacer regions and reported the presence of $tRNA^{Ile}$ and $tRNA^{Ala}$ genes in it. However, both tRNA genes are split by very long intervening sequences which exhibit a certain degree of homology.

In this work, we present the primary structure of soybean spacer rDNA, its complete nucleotide sequence with the presence of $tRNA^{Ile}$ and $tRNA^{Ala}$ genes, which contain intervening sequences, different in length from those found in maize and tobacco chloroplast DNA.

Material and methods

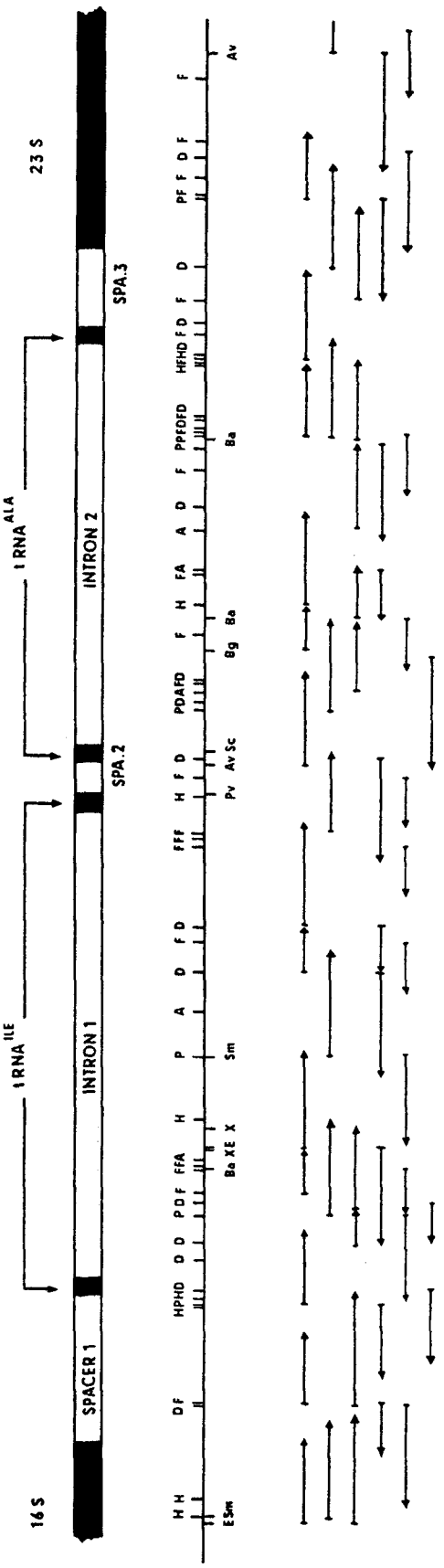
Materials

Restriction endonucleases Ava I, Ava II, Bam HI, Bgl II, Dde I, Hin FI, Hae III, Hpa II, Pvu II, Sac I, Sma I and Xho I were purchased from Bethesda Research Labs., Inc.

Restriction endonuclease mapping

Cloned soybean cpDNA fragments pDS10 (3.9 kb, Sac I) and pBP9 (4–1 kb, Pvu II) containing 16S and 23S rRNA genes respectively were employed for restriction endonuclease map-

* To whom correspondence should be addressed



ping with various enzymes. The cleavage sites were ordered by single and double digestions as previously described in Singh et al. (1984). Figure 1 shows the positioning of cleavage sites of selected restriction endonucleases.

Radioactive labeling, filters and hybridization

Labeling of fragments, transfers to filters and hybridization were done as described previously (Singh et al. 1984).

DNA sequencing

Recombinant plasmids pDS10 and pBP9 DNA were digested completely with various restriction enzymes and the digests resolved by polyacrylamide gel electrophoresis. DNA sequence analysis was performed according to the procedures described by Maxam and Gilbert (1977).

Results and discussion

Sequencing of the 16S–23S spacer from soybean (Glycine max L.) chloroplasts

The soybean chloroplast recombinant plasmids pDS10 (3.9 kb, Sac I) and pBP9 (4.1 kb, Pvu II) containing the 16S–23S spacer region were isolated and used for sequencing by the method of Maxam and Gilbert (1977). Thirteen different enzymes (Ava II, Dde I, Hin FI, Hae III, Hpa II, Eco RI, Ava I, Bam HI, Bgl II, Sma I, Xho I, Pvu II and Sac I) were used for fine mapping and sequencing according to the strategy presented in Fig. 1. The 16S–23S spacer is over 3000 bp long and its nucleotide sequence (RNA-like strand) is shown in Fig. 2. To our knowledge, this is the first sequence analysis of a complete 16S–23S rDNA spacer from a dicotyledon leguminoseae plant. The spacer sequence is confined at its 5′ side by the 3′ terminus of the 16S rRNA gene and at its 3′ side by the 5′ terminus of the 23S rRNA gene.

16S–23S spacer from soybean chloroplasts contains two tRNA genes

Screening for GTTC sequences which are indicative of the GT4C sequences commonly present in the T loop of the tRNA species (Springer et al. 1981), we could find two tRNA genes which are interrupted by long intervening sequences. Based on their anticodon sequences, they are identified to be $tRNA^{Ile}$ and $tRNA^{Ala}$ genes. Quite similar to other spacer regions in maize (Koch et al. 1981) and tobacco (Takaira and Sugiura 1982), the coding regions of genes for $tRNA^{Ile}$ and

Fig. 1. A: Ava II, D: Dde I, F: HinF I, H: Hae III, P: Hpa II, E: EcoR I, Av: Ava I, Ba: Bam HI, Bg: Bgl II, Sm: Sma I, X: Xho I, Pv: Pvu II, Sc: Sac I. (SPA 2: spacer 2), (SPA 3: spacer 3)

$tRNA^{Ala}$ are split by intervening sequences of 953 bp (intron I) and 811 bp (intron II) respectively. These introns are located between positions 506 and 1459 and between positions 1596 and 2407 respectively (Fig. 2).

Homology/comparison of 16S–23S spacer tRNA sequences

Table 1 presents a quantitative comparison of the soybean chloroplast 16S–23S spacer tRNA sequences with higher plants (maize and tobacco), algae (*Euglena* and *Anacystis nidulans*) and bacteria (*E. coli* and *Bacillus* S.). The degree of homology was highest for the $tRNA^{Ile}$ between soybean, maize and tobacco (100%), while strong homology was observed with *Euglena* and *Anacystis,* followed by much lower homology with *E. coli* and *Bacillus* S. When one compared the $tRNA^{Ala}$, there was highest homology with soybean and maize (100%), on the other hand, tobacco showed slightly less homology at 98.6%. Among the algae, *Euglena* was stronger than *Anacystis*, which was quite opposite to what we observed for $tRNA^{Ile}$. There was no significant difference between the *E. coli* and *Bacillus* S. In addition to the sequence homology, positional homology exists for the two chloroplast tRNA genes, since they lie in the same order between the 16S and 23S rRNA genes, as in maize, tobacco, *Euglena, Anacystis, E. coli,* and *Bacillus* S.

Table 1. rRNAs comparison; degree of homology of $rDNA^{Ile}$ and $rDNA^{Ala}$ in several species (the CCA ends or their equivalent are not included in the comparison)

Ref	tDNA Ala \ tDNA Ile	soybean	maize	tobacco	*Euglena g.*	*Anacystis n.*	*Escherichia c.*	*Bacillus s.*
	soybean		100	100	89.2	91.9	81.1	78.4
2	maize	100		100	89.2	91.9	81.1	78.4
3	tobacco	98.6	98.6		89.2	91.9	81.1	78.4
4	*Euglena g.*	91.9	91.9	91.9		85.1	77.0	72.9
5	*Anacystis n.*	87.8	87.8	87.8	89.2		86.5	83.8
6	*Escherichia c.*	81.1	81.1	82.4	82.4	83.8		90.1
7	*Bacillus s.*	77.0	77.0	78.4	77.0	86.5	91.9	

% of homology

% of homology

Comparison of the 16S–23S spacer regions in several species

Recently, a number of reports have been published describing the nucleotide sequences of chloroplast 16S–23S spacer and the presence of two split tRNA genes in maize (Koch et al. 1981) and tobacco (Takaira and Sugiura 1982). The apparent procaryotic nature of the spacer tRNA genes and of the entire soybean, maize and tobacco chloroplast rRNA operon is challenged by the presence of two intervening sequences. The soybean $tRNA^{Ile}$ and $tRNA^{Ala}$ genes contain 953 and 811 bp intervening sequences respectively. A comparison of the spacer regions in several species is presented in Table 2. These spacer sequences for maize, tobacco and soybean are similar to each other. The soybean spacer sequence (2433 bp) is longer than maize and tobacco counterparts by 25 and 353 bp respectively. This is probably due to a 253 bp deletion in tobacco intron I and a 113 bp deletion in intron II compared to the soybean.

Table 2. Size comparisons of the 16S–23S spacer region in several species; references are as in Fig. 2; sizes are given in base pairs at the DNA level

Species \ Region	Spa 1	Ile 1	Int 1	Ile 2	Spa 2	Ala 1	Int 2	Ala 2	Spa 3	Total
soybean	303	36	953	36	64	36	811	37	157	2,433
tobacco	300	36	707	36	64	36	710	37	154	2,080
maize	301	36	949	36	64	36	806	37	143	2,408
Euglena g.	87		73		9		73		16	258
Anacystis n.	151		74		33		73		201	532
Escherichia c.	68		74		42		73		174	431
Bacillus s.	102		74		11		73		83	341

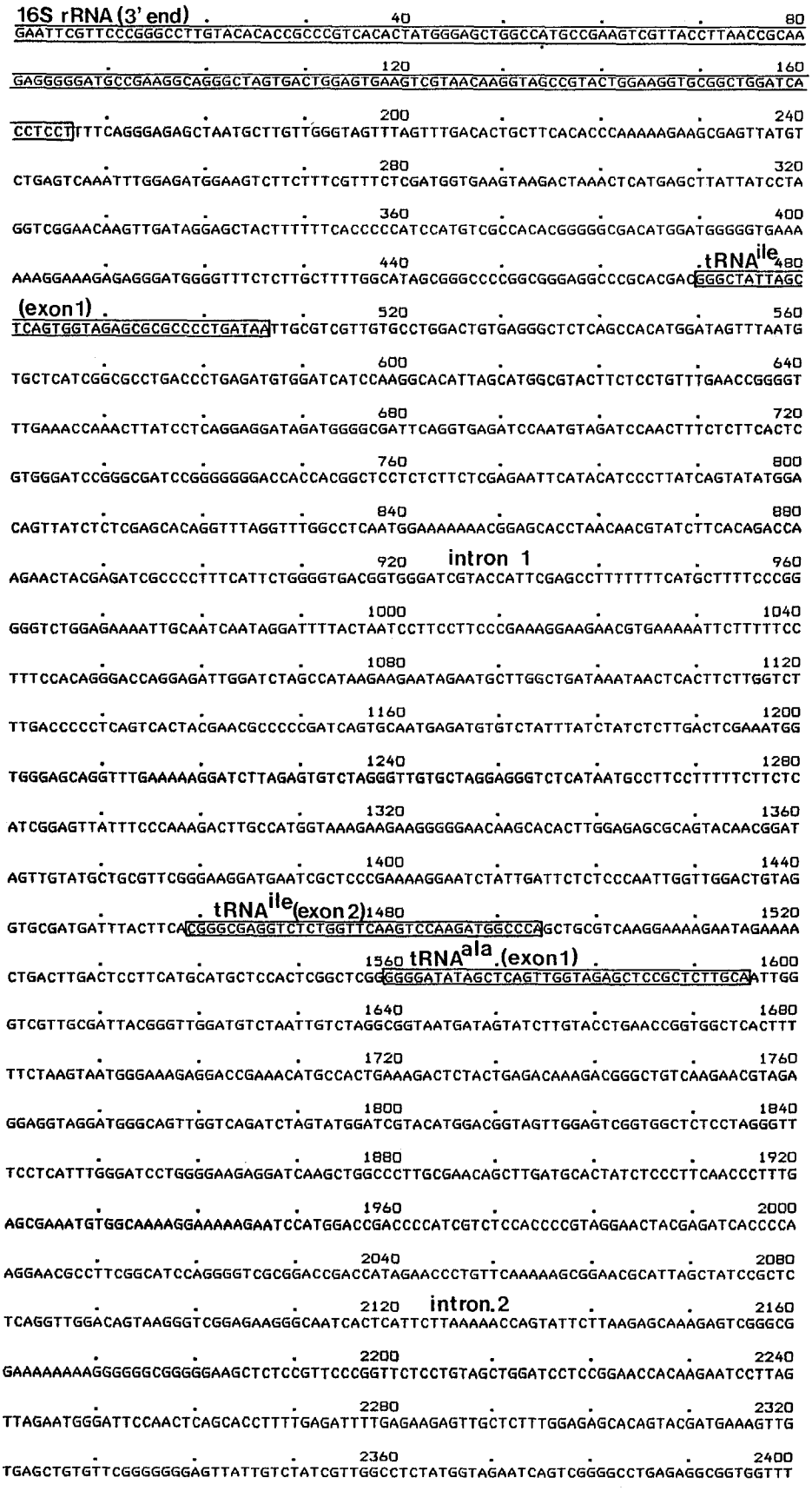


Fig. 2. Nucleotide sequence of the 16S – 23S spacer rDNA from soybean (*Glycine max*) chloroplasts

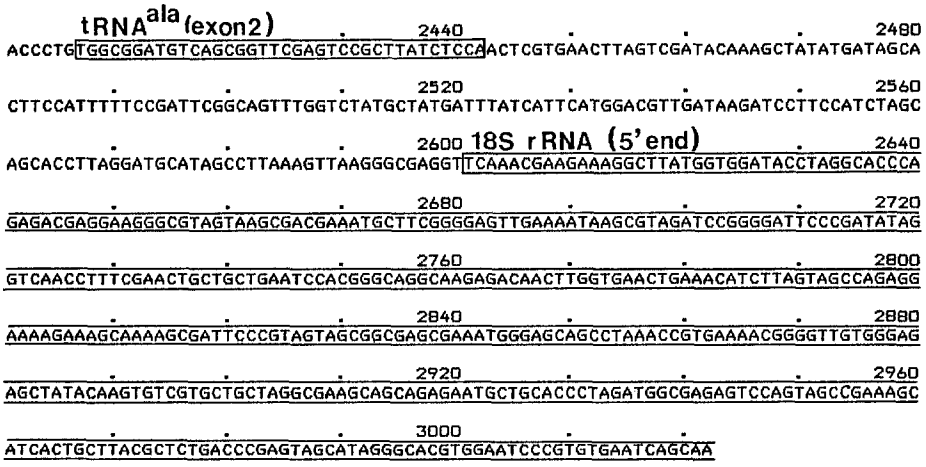


Fig. 2. (continued)

Table 3. Degree of homology between soybean, tobacco and maize for the 16S–23S spacer region

Region considered	Species compared % of homology		
	soy/tob	soy/maize	tob/maize
3'-16S (166nt)	97.7	92.4	91.2
spacer	86.8	77.9	82.8
$tDNA^{Ile}$	100	100	100
intron 1*	93.6	90.1	91.4
$tDNA^{Ala}$	98.6	100	98.6
intron 2**	95.0	90.0	92.8
5'-23S (422nt)	93.7	92.8	92.1

* Comparison for intron I does not take into account 253 nt region absent in tobacco when compared to soybean in primary structure alignment (Since this region is present in maize, the comparison between soybean and maize does include this region)

** Same remark as in * for the 113 nt region absent in intron 2 of tobacco when compared to soybean

The genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ are 72 and 73 bp long respectively, which are split by the presence of two introns which vary in size. Figure 2 shows the 16S–23S spacer region divided into spacer 1, 2 and 3, $tRNA^{Ile}$ 1 and 2, $tRNA^{Ala}$ 1 and 2 and introns I and II. While the spacer regions 1 and 3 are almost similar in size, the size of the two introns vary. Compared to the soybean introns I and II, the tobacco introns I and II are shorter by 246 and 101 bp, whereas the maize introns I and II are shorter by 4 and 5 bp only. Among the procaryotes listed in Fig. 2, the two tRNA genes for Ile and Ala are 74 and 73 bp long. However, the size differences between spacer regions 1, 2 and 3 show considerable variation, without the presence of introns.

Table 3 shows the degree of homology between soybean, tobacco and maize for the 16S–23S spacer region. The 3'-16S region, which is 166 bp long, is 97.7% homologous to the tobacco region, compared to maize at 91.2%. It is clear from the data that the spacer, intron I, intron II and 5'-23S (422 bp long) region all show greater homology between soybean/tobacco than between soybean/maize or tobacco/maize. For $tRNA^{Ile}$ sequence, there is 100% homology among all species, while $tRNA^{Ala}$ sequence is between 98.6–100% homology. It should be pointed out that in this analysis, the 253 and 113 bp regions absent in intron I and II respectively, are not taken into account.

Homology between two introns

A comparison of the two soybean intron sequences revealed remarkable sequence similarity. A similar observation has been made in the maize and tobacco introns, with many common sequences in corresponding positions. For example, (5') ATT–GTCGTTG YG was found at 5' end, YARGAACTACGAGAT-RCCCC in the middle and GCGRTGRTTTACYYYR (3') at the 3' end of each intron. Further homologies can be observed between the two introns, though the length and spacing of identical sequences varies considerably. These observations indicate that the homology appears strong enough to suggest that these introns in chloroplast spacer tRNA genes are of common origin and that their sequences and sizes have diverged gradually. As discussed above, it is possible that the conserved sequences at the intron borders represent signal structures recognized by the splicing enzymes. It is possible that such recognition may be of either primary or secondary structure. Lack of similar homology in the intervening sequences from yeast tRNA genes does not support this possibility (Abelson 1979; Knapp et al. 1979). Koch et al. (1981) suggest that the two intron sequences may be interpreted as either functional or

vestigial transposable elements (Carlos and Miller 1980). Formation of short direct repeats from the target sequences positioned at either side of a transposable element are generally a consequence of their insertion (Carlos and Miller 1980).

A comparison of the tobacco and maize spacers, with those of *E. gracilis* chloroplasts and *E. coli rrnD* operon (Takaira and Sugiura 1982), indicated that the coding regions for $tRNA^{Ile}$ and $tRNA^{Ala}$ are highly conserved, in spite of remarkable divergency in intergenic regions.

The tobacco $tRNA^{Ile}$ sequence showed 83% and 81% homology, and the $tRNA^{Ala}$ sequence showed 90% and 78% homology, with those of *E. gracilis* chloroplasts and *E. coli rrnD*, respectively. The 3′ terminal CCA sequence is not coded for by the chloroplast DNA like other chloroplast tRNA genes sequenced (Koch et al. 1981; Graf et al. 1980; Orozco et al. 1980; Kato et al. 1981; Tohdoh et al. 1981; Schwarz et al. 1981 a, b).

In soybean, we have found two open reading frames (ORFs) of 65 and 101 amino acids in intron I and 66 and 50 amino acids in intron II. Koch et al. (1981) have pointed out that the maize introns have possible reading frames for 123 and 45 amino acids long respectively. However, such long reading frames could not be detected in corresponding regions (starting at positions 443 and 1513) of the tobacco introns. Further analysis has to be carried out to determine the nature and meaning of these ORFs. Earlier, we have stated that the intron size differences between soybean and tobacco are mainly due to two regions absent in tobacco (253 bp for intron I and 101 bp for intron II). It would be of interest to determine whether these regions are deletions (in tobacco) or insertions (in soybean) for the evolution of these intervening sequences and also since these regions belong in the soybean to the ORFs mentioned above.

Acknowledgements. This work was supported by a Grant A-1984 from the Natural Sciences and Engineering Research Council of Canada.

References

Abelson J (1979) RNA processing and the intervening sequence problem. Ann Rev Biochem 48: 1035–1069

Bedbrook JR, Kolodner R (1979) The structure of chloroplast DNA. Ann Rev Plant Physiol 30: 593–620

Bohnert HR, Driesel AJ, Crouse EJ, Gordon K, Herrmann RG, Steinmetz AA, Mubumbila M, Keller M, Burkard G, Weil JH (1979) Presence of a transfer RNA gene in the spacer sequence between the 16S and 23S rRNA genes of spinach chloroplast DNA. FEBS Lett 103: 52–56

Carlos MP, Miller JH (1980) Transposable elements. Cell 20: 579–595

Graf L, Kossel H, Stutz E (1980) Sequencing of 16S–23S spacer in a ribosomal RNA operon of *Euglena gracilis* chloroplast DNA reveals two tRNA genes. Nature 286: 908–910

Kato A, Shimada H, Kusuda M, Sugiura M (1981) The nucleotide sequences of two $tRNA_{ASN}$ genes from tobacco chloroplast. Nucl Acids Res 9: 5601–5607

Knapp G, Ogden RC, Peebles CL, Abelson J (1979) Splicing of yeast tRNA precursors: Structure of the reaction intermediates. Cell 18: 37–45

Koch W, Edwards K, Kossel H (1981) Sequencing of the 16S–23S spacer in a ribosomal RNA operon of *Zea mays* chloroplast DNA. Cell 25: 203–213

Malnoe P, Rochaix JD (1978) Localization of 4S RNA genes on the chloroplast genome of *Chlamydomonas reinhardii.* Mol Gen Genet 166: 269–275

Maxam A, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74: 560–564

Morgan EA, Ikemura T, Nomura M (1977) Identification of spacer tRNA genes in individual ribosomal RNA transcription units of *Escherichia coli.* Proc Natl Acad Sci USA 74: 2710–2714

Orozco EM Jr, Rushlow KE, Dodd JR, Hallick RB (1980) *Euglena gracilis* ribosomal RNA transcription units II: Nucleotide sequence homology between the 16S–23S ribosomal RNA spacer and the 16S ribosomal RNA leader regions. J Biol Chem 255: 10997–11003

Schwarz Z, Jolly SO, Steinmetz AA, Bogorad L (1981a) Overlapping divergent genes in the maize chloroplast chromosome and in vitro transcription of the gene for $tRNA^{HIS}$. Proc Natl Acad Sci USA 78: 3423–3427

Schwarz Z, Kossel H, Schwarz E, Bogorad L (1981b) A gene coding for $tRNA^{Val}$ is located near 5′ terminus of 16S rRNA gene in *Zea mays* chloroplast genome. Proc Natl Acad Sci USA 78: 4748–4752

Sekiya T, Nishimura S (1979) Sequence of the gene for isoleucine $tRNA_1$ and the surrounding region in a ribosomal RNA operon of *Escherichia coli.* Nucl Acids Res 6: 575–592

Singh G, Lanversin G de, Straus NA, Pillay DTN (1984) Characterization of soybean chloroplast ribosomal RNA genes. Curr Gen 9: 99–105

Springl M, Grueter F, Spetzlaus A, Gauss DH (1981) Compilation of tRNA sequences. Nucl Acids Res 9: r1–r23

Takaira F, Sugiura M (1982) Nucleotide sequence of the 16S–23S spacer region in an rRNA gene cluster from tobacco chloroplast DNA. Nucl Acids Res 10: 2665–2676

Tohdoh N, Shinozaki K, Sugiura M (1981) Sequence of a putative promoter region for the rRNA genes of tobacco chloroplast DNA. Nucl Acids Res 9: 5399–5406

Young RA, Macklis R, Steitz JA (1979) Sequence of the 16S–23S spacer region in two ribosomal RNA operons of *Escherichia coli.* J Biol Chem 254: 3264–3271